DOCUMENT RESUME

ED 287 153 CS 008 923

AUTHOR Sherman, Thomas F.; Merschman, Jane
TITLE Reading Diagnostic Tools: Review of Woodcock Reading Mastery Test.
PUB DATE Sep 87
NOTE 15p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Book/Product Reviews (072)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Criterion Referenced Tests; *Diagnostic Tests; Elementary Secondary Education; *Reading Diagnosis; Reading Skills; *Reading Tests; Screening Tests; *Test Construction; Test Items; Test Reliability; Test Reviews; Test Use; Test Validity
IDENTIFIERS *Woodcock Reading Mastery Test

ABSTRACT

The development and assessment of the Woodcock Reading Mastery Test (WRMT) is described in this paper. The first section, after a brief description of the test, outlines the development of the test, including its purpose, how it was tested and calibrated, its administration and scoring, use and interpretation of scores obtained, and statistical development. The second section of the paper offers assessments of the test from a dozen reviewers, five of whom thought the test was not useful or seriously flawed due to, among other things, (1) a lack of a description in the manual for how the test can be used diagnostically, (2) sex-role stereotyping in the test, and (3) lack of evidence of attempts to measure inference, logic, or analysis. The positive reviews in this section rated the test as useful in that, among other things, it offers a wide variety of interpretive scores, presents directions and interpretations of scores clearly in the manual, and makes alternate forms of the test available. Concluding remarks suggest that the WRMT is not what the author promised it would be, although it may have value as an initial screening test. (Seven references are included.) (JC)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED287153

# READING DIAGNOSTIC TOOLS: REVIEW OF WOODCOCK READING MASTERY TEST

BY

THOMAS F. SHERMAN

AND

JANE MERSCHMAN

SEPTEMBER 1987

BEMIDJI STATE UNIVERSITY READING CLINIC

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Thomas F. Sherman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS008923



BEST COPY AVAILABLE

SYNOPSIS:

The Woodcock Reading Mastery Test (WRMT) is an individually administered, untimed test for students in kindergarten through the twelfth grade. The WRMT consists of five subtests: letter identification, word identification, word attack, word comprehension, and passage comprehension. The WRMT was last published in 1973 and i available from the American Guidance Service, Incorporated. The approximate administration time is thirty minutes. Two forms are available (A & B) with socioeconomic-status adjusted norms, if needed. The WRMT consists of a manual, an easel containing the 400 test times, and a packet of student response forms.

PURPOSE:

Richard W. Woodcock designed and developed the WRMT to fill a need of reading diagnosticians. The three-fold objective of this instrument is to measure skill in each subtest area with greater precision than is available from other tests; the administration of the tests should be as simple as possible to learn and procedures should be as simple as possible to administer; and new ways of interpreting test scores should be incorporated, which would allow more useful interpretations of the subject's status.

DESCRIPTION AND REACTION TO COMPONENTS OF THE INSTRUMENT

The Manual includes the following descriptions: development of the test, administration and scoring, interpretation of scores and how they should be used, and statistics. A description and reaction to each of these areas will follow.

Development: The traditional procedures of test development were used, but also two new measurements were used (the Rasch-Wright item-analysis and the

principle of matrix-sampling). The WRMT had its origin in the Beginning Reading Test by Woodcock and Pfost (1967). This test had been developed to meet the need for a highly reliable measure of reading which would discriminate among pupils achieving below second grade. The original concept was extended from first grade up through grade twelve. To meet the objectives, it was decided to use the open-ended format (to minimize guessing) and to individually administer the test. During the test item preparation stage, the goal was to develop and evaluate at least twice as many items as would be needed to construct the two projected forms. The initial pool contained 2,417 items. Next, came the calibration testing. A total of 36,527 calibration tests were individually administered. During this stage, the opportunity was taken to establish the clearest possible set of directions for the WRMT. The "item analysis" stage was conducted by Rasch and Wright. Their procedures provided a more thorough evaluation of item performance than did a traditional method. At the end of this editing stage, 1,332 items were considered as having adequately met the criteria of fitting the Rasch-Wright model.

The next step was to chain the difficulty values from each calibration test. Norming the scale was the next procedure implemented. One advantage of having used the Rasch-Wright calibration is that once a set of test items has been calibrated, the test may be normed by using a subset of the items. The norming tests contained a total of 75 test items. In general, there was approximately a 10-point difficulty difference (mastery scale) between each item in the norming tests. The norming of the WRMT took place over two years. 5000 subjects were tested in grades kindergarten through twelve. The nature of the sample was that of a "stratified random". Before calculating grade and age equivalency scores and percentile ranks,

normative data was weighted to conform to the U.S. proportions. The data was collected from students enrolled in regular classrooms. The reliability of the WRMT is discussed according to three kinds of information. First, the test's split-half reliabilities (adequate), next the test-retest alternate-form reliabilities (adequate), and finally the standard errors of measurement (between two and four points on the Mastery Scale). The validity of the WRMT is discussed according to three kinds of information. First, the content validity (adequate), the multimethod-multitrait matrix (ranges from .06 to .97), and the predictive validity (ranges from .30 to .80).

Administration and Scoring: The manual gives specific instructions for the administration of the WRMT. It mentions the qualifications needed by the examiner, the physical setting, establishment of rapport with the examinee, and finally discusses general instructions, each subtest is also treated in an individual manner. The manual specifically identifies supplementary information and specific suggestions regarding each subtest. The scoring descriptions are clear and examples are clearly given. A basal of five consecutive correct responses and a ceiling of five consecutive incorrect responses are used as a foundation for scoring. The examinee's estimated reading level is used as a starting point to determine the basal response level. A table on the easel indicates the question number that the administrator should begin with. Instructors are included for determining the basal and the ceiling should the child not perform at the appropriate level.

Use and Interpretation of Scores: Six type of scores are derived from the computation of the raw score. These include the easy reading level, the reading grade score, the failure reading level, the relative mastery score, the percentile rank, and the normal curve equivalent. Tables are provided for ease of these computations. Also, a computer disk is available in which the raw scores are entered and the other scores are figured by the computer. The author explains each of the scores in detail and how they should be interpreted.

Statistical Development: The norming population included 5000 subjects in grades kindergarten through twelve over a two-year period. Various statistical analyses were employed to provide reliability and validity data for the WRMT.

The Easel-Kit contains the four hundred items consecutively numbered which the children will be tested on. Samples are only given in two of the five subtests. The Easel-Kit is specially designed for representing and storing the test materials adequately. When opened, the Easel-Kit takes an easel shape which allows presentation of the test items to the subject while at the same time providing the examiner with instructions, a copy of the items, and a key to acceptable items on the other side.

The score sheets are simply and compactly designed. They are easy to follow and record responses on. The last page of the score sheet also includes the Mastery Profile, which portrays the examinee's performance on the test in terms of "instructional range" and a predicted percent of mastery at significant points along a grade scale.

Comments from the Mental Measurements Yearbook: The Woodcock Reading Mastery Test was reviewed by two different individuals in The Eighth Mental Measurement Yearbook. Also included were four reviews on the WRMT from various professional journals.

The first reviewer is Carol Anne Dwyer, Program Director of Elementary and Secondary School Programs, Educational Testing Service at Princeton, New Jersey. She begins her review by stating that the author's promise of a solution to problems in reading assessment is unfulfilled in this test. She feels that the primary objective (to provide precise measures of reading ability which are easy to administer and interpret) is only partially achieved through the WRMT. But, she feels that a useful feature is the coverage of grades kindergarten through twelve in a single instrument. Another plus for the WRMT is that the tests, graphics and overall design are attractive. Although, she does feel that the artwork is a bit old-fashioned.

Dealing with the subtests separately, she feels that the Letter Identification Subtest should be eliminated completely. She feels that it would be better utilized in a readiness test. Next, the Word Identification Test may offend those professionals who stress word identification through context clues. Although, she does feel it is a well-done example of a traditional reading task. Third, she feels that the Word Attack Subtest is another well-done example of a traditional reading task in its use of nonsense words. The Word Comprehension Subtest is in large part vocabulary. The reviewer feels that the examinee will also need distinct reasoning and classification skills. Finally, the Passage

Comprehension Subtest consists of a modified cloze (because the deletions are not arrived at by specific rules) which the reviewer feels the author used to purposefully omit key words and phrases.

The reviewer goes on further to criticize the WRMT by stating that she sees no evidence of attempts to measure inference, logic, or analysis. She feels that the test context seems best suited for a global screening measure for reading disability and not for any kind of a precise decision. Finally, she feels that sex-roles are stereotyped in this test.
The reviewer feels that the manual combines administrative instructions and technical data well. Although, she feels that it could be overwhelming to someone who is just looking for directions.

The reviewer feels that the administration of the WRMT is relatively simple. She also feels that the easel format is convenient and sturdy. Another advantage that the reviewer mentions is the short administration time. Negatively, she feels that the WRMT is difficult to score and interpret.

Ms. Dwyer mentions that the test items were analyzed and calibrated using Rasch-Wright procedures. However, she does not feel that enough information is given to determine if the Rasch-Wright models assumptions were met. She also feels that although the criterion-referenced testing was done, the test is clearly norm-referenced because meaningful criterion-referenced interpretations weren't provided.

When addressing the norms, the reviewer feels that they are clearly presented, thoroughly researched, and well-constructed. However, she feels that the rationale for separate sex and SES-adjusted norms is weak. Who would use these norms and why? The reviewer feels that the author has given us no guidelines for when and how these norms should be used. Ms. Dwyer speaks about the reliability and is not quite happy. She notes that the split-half and alternative-form reliabilities are only reported for grades 2 and 7. Also, the author has included pretest reliability data which the reviewer feels is misleading because the pretests are not identical to the final forms. Finally, she notes the publisher's catalog claims split-half reliabilities in the .90-.99 range. In actuality, the reviewer found a range of .02-.99, which she feels are adequate but not exceptional.

In conclusion, Ms. Dwyer feels that the WRMT is seriously flawed and the claims that are made by the author are not supported by data.

The second reviewer of the WRMT was J. Jaap Tuinman, Professor of Education at Simon Fraser University in Burnaby British Columbia, Canada. In opening, this reviewer states that this is the most unusual battery of tests in the decade and somewhat misleading. He feels that there is not support for the criterion-referencing of the test. Finally, he notes that there are no traditional grade scores used.

When dealing with each subtest individually, the reviewer felt that the Letter Identification Subtest was unusual and useless. The word Identification Subtest tended to measure different functions for different

children. Next, the Word Attack Subtest should have had error analysis rather than an overall score. The reviewer also criticizes the use of real words after the child has been told that these words are all nonsense words. In regard to the Word Comprehension Subtest, the reviewer feels that the subtest measures reasoning more than it measures comprehension. He feels that poor readers are penalized. Finally, the Passage Comprehension Subtest penalizes the poor reader, according to the reviewer.

When addressing content validity, the reviewer feels that a problem with the rest is the use of pictures in only 29% of the items. He feels they should be used throughout, or not at all. He also sights research on the cloze technique and how it is largely a measure of local redundancy and that it fails to measure understanding of large idea units.

Professor Tuinman addresses the technical data next. Generally speaking, he feels that the test is more reliable in the lower grades. He sees no validity studies involving external criteria, which he sees as a weakness. He also feels that the author's claim to provide the user with a set of criterion-referenced scores is not met.

In summary, the Professor feels that the WRMT has a number of strong points. First, it has a wide variety of interpretive scores. Also, the manual is clear and concise. The test directions are fairly simple. Finally, there are not multiple-choice questions.

However, he feels that the WRMT has a number of weaknesses. First, the stated administrative time is unrealistic for a poor reader. Also, the

author's choice of the 90% success rate as mastery is questionable. Finally, the criterion-referenced interpretation claims are largely unfounded.

Professor Tuinman feels that this test can be a valuable tool when used by an experienced reading diagnostician and should not be used by the general population.

Alex Bannatyne (from the August-September issue of the Journal of Learning Disabilities, 1974) feels that the most innovative feature of the WRMT is the inclusion of the SES (socioeconomic status) adjusted norms. He also felt that the Mastery tests would be useful for clinical or research purposes. Negatively, he did not see that rate of reading assessed, or the question of syntax covered. In summary, he felt it would be a valuable addition to a diagnostician's assessment battery as it is easy to administer and score.

Richard L. Allington (1976) feels that rarely is a test developed that offers a variety of unique features while maintaining or improving assessment effectiveness and efficiency. He was impressed with the validity and reliability data provided. He also felt that the use of the Rasch-Wright analysis procedures and the development of a criterion-referenced Mastery Scale are unique features of the WRMT. The reviewer uses the instrument for a year. After that time he found the WRMT most useful for assessing reading achievement. He found the manual clearly presents directions and interpretation of test results. He also liked the availability of alternate forms. In summary, he felt that with experience

in administration of the WRMT, he can support the claims made by the author and feels it is an excellent individual reading achievement test.

Cherry Houck and Larry A. Harris (1976) began by stating that they feel that there is insufficient data to support the external validity of the WRMT and the content validity is open to question. They agree with the first two individual reviewers in the area of the separate subtests. They feel that the WRMT is easy to administer and rapid to complete, which they feel allows the examinee to sustain its best effort. They feel that exposure to only a few items on every page should decrease any frustration felt. They feel that there are three concepts in this test that represent newer interpretation procedures: Relative Mastery, Achievement Index, and Relative Mastery at grade level. They felt that in terms of SES, it was more time consuming than is desirable. They also felt that the norm-referenced and criterion-referenced scales were overrated. In summary, they state that criterion-referenced procedures are not included (as is stated in the manual), more evidence is needed to support external validity, and they feel that the WRMT can effectively serve as a screening device.

Barton B. Proger (1975) feels that the WRMT is the only formal instrument that has built in criterion-referenced measurement and norm-referenced measurement. This reviewer questions whether consumers can take advantage of all of the options. He also feels that the manual might be a bit overwhelming to an ordinary test consumer. One of his criticisms is that not every subtest has examples and he feels that this could adversely affect the child who does not understand verbal directions. He also feels

that the validity data is sparse. He states that the Easel-Kit is suitable, but both of the forms could have been in one binder. In conclusion, he feels that careful deployment of the Rasch-Wright model is noteworthy, along with the inclusion of the Criterion-referenced measurements and the norm-referenced measurements.

Comments from other Sources:

David Memory, Glen Powell and Byron Callaway (1980) conducted a study of the assessment characteristics of the WRMT. In their study, they compared scores from WRMT form A with information from Spache Diagnostic Reading Scales and the Slosson Oral Reading Test. They used 62 children in their study. They found that it is convenient for diagnosing strengths and weaknesses. Listed under the attractive features are: Design of the Word Identification Test to predict 96% accuracy in word recognition, and the design of the Passage Comprehension test to predict 75% accuracy in comprehension. In conclusion, they feel that the WRMT seems to be valid for assessing reading levels of students. Also, comprehension is best assessed by the Passage Comprehension Test. Finally, the Word Identification Test is (on average) one year lower than the Grade Equivalency scores on sight vocabulary words in the two other tests.

James L. Laffey and Donna Kelly (1979) reviewed the WRMT and came to the following conclusions. They agree with the first two reviewers on the subject of the separate subtests. One major fault of the manual is that it does not describe how the test can be used diagnostically. They also caution the administrator against interpreting the total score. They question some of the variations on the conversion scales. In conclusion,

they feel that the WRMT is not a good diagnostic test and should not be used as one. They question the viability of a test that spans the grades of kindergarten through twelve. Finally, they feel that the total score is inflated.

Conclusion: The WRMT does not appear to be what the author has promised that it would be. The instrument value would seem to be as an initial screening test. The norms and the appropriateness of some of the subtests are questionable. The strengths of the WRMT are the ease of administration, the short administration time, the open-ended questions, the clear printing, the absence of clutter on the pages, the minimal scoring time especially with the computer scoring program, the variety of reporting scores which include percentiles, and the two forms for pre and post measures.

REFERENCES

Allington, Richard L. "Review of the Woodcock Reading Mastery Tests." Journal of Reading, November 1976.

Bannatyne, Alex. "Review of the Woodcock Reading Mastery Tests." The Journal of Learning Disabilities, August-September, 1974.

Dwyer, Carol A., & Tuinman, J. Jaap (1979). Woodcock Reading Mastery Tests. The eighth mental measurements yearbook (pp. 1303-1308). Lincoln, Nebraska: Buros Institute of Mental Measurements of the University of Nebraska.

Houck, Cherry, & Harris, Larry A. "Review of the Woodcock Reading Mastery Tests." The Journal of School Psychology, Spring, 1976.

Laffey, James L., & Kelly, Donna. "Test review of the Woodcock Reading Mastery Tests." Reading Teacher, December, 1979.

Memory, David & Others. "A study of the Assessment Characteristics of 'Woodcock Reading Mastery Tests'." Reading Improvement, Spring, 1980.

Proger, Barton B. "Review of the Woodcock Reading Mastery Tests." Journal of Special Education, Winter, 1975.